

## *NotI* Genomic Cleavage Map of *Escherichia coli* K-12 Strain MG1655

JOE DON HEATH, JEFFREY D. PERKINS, BALDEV SHARMA,  
AND GEORGE M. WEINSTOCK\*

*Department of Biochemistry and Molecular Biology, University of Texas Medical School,  
University of Texas Health Science Center at Houston, 6431 Fannin Street,  
P.O. Box 20708, Houston, Texas 77225*

Received 26 August 1991/Accepted 2 November 1991

Several approaches were used to construct a complete *NotI* restriction enzyme cleavage map of the genome of *Escherichia coli* MG1655. The approaches included use of transposable element insertions that created auxotrophic mutations and introduced a *NotI* site into the genome, hybridization of *NotI* fragments to the ordered  $\lambda$  library constructed by Kohara et al. (BioTechniques 10:474–477, 1991), Southern blotting of *NotI* digests with cloned genes as probes, and analysis of the known *E. coli* DNA sequence for *NotI* sites. In all, 22 *NotI* cleavage sites were mapped along with 26 transposon insertions. These sites were localized to clones in the  $\lambda$  library and, when possible, sequenced genes. The map was compared with that of strain EMG2, a wild-type *E. coli* K-12 strain, and several differences were found, including a region of about 600 kb with an altered restriction pattern and an additional fragment in MG1655. Comparison of MG1655 with other strains revealed minor differences but indicated that this map was representative of that for many commonly used *E. coli* K-12 strains.

Over the last few years, a number of major resources for genetic and physical analysis of the *Escherichia coli* K-12 chromosome have been developed. One of the first advances was the construction of the *NotI* restriction enzyme cleavage map of the genome of strain EMG2 (27), a strain considered to be closely related to the original wild-type *E. coli* K-12 isolate. Kohara and coworkers (20) constructed and ordered a set of phage  $\lambda$  clones with nearly complete coverage of the genome. DNA from a subset of these phages (the miniset) is available as an ordered set on a single filter that can be used in hybridization experiments to localize sequences within the chromosome (21). A nearly complete cosmid library has also been constructed (28). A collection of transposon insertions has been assembled containing a marker roughly every minute around the chromosome (26). Finally, the total published sequence of the *E. coli* chromosome, representing approximately 35% of the genome, has been compiled and made available along with software for analysis (22, 23).

These powerful resources provide new tools for the analysis of *E. coli*. One potential complication is that different strains of *E. coli* K-12 have been used for these projects. For example, the *NotI* map used strain EMG2 (9), which, although it is considered to be closely related to the original wild-type *E. coli* K-12 isolated in 1922, is not itself widely used for studies of *E. coli*. The ordered phage  $\lambda$  and cosmid libraries used strain W3110, a popular strain that nevertheless contains a large inversion in its chromosome (17) as well as other changes (22–24, 28). The transposon collection is derived from many sources but relies on strain MG1655 (16), another popular strain, as a common carrier. Finally, the DNA sequence data base is drawn from many strains. It is not clear how much of a problem this lack of isogenicity will pose in applying these resources. However, it is clear that there are differences in the genomes of these strains, and we report on some of these here.

Because much of our own work relies on strain MG1655, we have determined the *NotI* restriction enzyme cleavage map for this strain and present it here. The map differs in several regions from that of EMG2. We also present a correlation between this map and the phage library of Kohara et al. and the *E. coli* sequence data base available from the National Center for Biotechnology Information. Examination of other commonly used *E. coli* strains shows that the MG1655 map is representative, although differences are observed.

### MATERIALS AND METHODS

**Bacterial strains.** All strains used in this study were derived from *E. coli* K-12. The auxotrophic strains used in this study were generated in this laboratory and in the Advanced Bacterial Genetics course at Cold Spring Harbor Laboratory. The transposition system used in generating the auxotrophic strains is described below. The strains with transposon insertions in *galETK*, *pyrD*, *pyrC*, *purF*, *purC*, *galP*, and *metE* were generated with the transposable element Tn10dCamNS. All remaining strains containing transposon insertions were generated by using the transposable element Tn10dCamMCS. The notable difference between these two elements is that Tn10dCamMCS contains the recognition sequences for *NotI*, *SfiI*, *XbaI*, *NheI*, and *DraI*, while Tn10dCamNS contains sites for *NotI* and *SfiI* only.

Table 1 lists the strains and plasmids used in this study. Strain EMG2, *E. coli* Genetic Stock Center number 4401, was obtained from B. Bachmann. Strain W1485 was provided by N. Sternberg, MG1655 was provided by K. B. Low, W3110 was provided by J. Lupski, AB1157 was provided by A. Ganesan, and MC4100 was provided by T. Silhavy.

**Media.** Minimal medium consisted of M63 salts, 0.2% glucose, 0.1 mg of thiamine per ml, and 1 mM MgSO<sub>4</sub>; rich medium was L broth (25). Final concentrations of antibiotics were as follows: ampicillin, 150  $\mu$ g/ml; spectinomycin, 50

\* Corresponding author.

TABLE 1. Bacteria and plasmids

Bacterium or plasmid	Description	Source or reference
<b>Bacteria</b>		
EMG2	F <sup>+</sup> $\lambda^+$	B. Bachmann (9)
W1485	F <sup>+</sup> $\lambda^-$	N. Sternberg (2)
MG1655	F <sup>-</sup> $\lambda^-$	K. B. Low (2)
W3110	F <sup>-</sup> $\lambda^-$ IN( <i>rrnD-rrnE</i> )I	J. Lupski (2)
AB1157	F <sup>-</sup> $\lambda^-$ <i>thr-1 ara-14 leuB6</i> $\Delta$ ( <i>gpt-proA</i> )62 <i>lacY1 tsx-33 supE44 galK2 Rac<sup>-</sup> hisG4(Oc) rfbD1 mgl-51 rpsL31 kdgK51 xyl-5 mtl-1 argE3 thi-1</i>	A. Ganesan (2)
MC4100	F <sup>-</sup> $\lambda^-$ <i>araD169</i> $\Delta$ ( <i>argF-lac</i> ) <i>U169 rpsL150 relA flbB5301 ptsF25 deoC1</i>	T. Silhavy (6)
GE1669	MG1655/pMS421/pMM12	This work
GE2515	MG1655/pMS421/pJD12	This work
<b>Plasmids</b>		
pMM12	<i>P<sub>lac</sub>-tnp</i> fusion, Tn10dCamNS element, <i>ori</i> pBR322	3
pMS421	<i>lacI<sup>q</sup></i> , <i>ori</i> pSC101	15
pJD12	<i>P<sub>lac</sub>-tnp</i> fusion, Tn10dCamMCS element, <i>ori</i> pBR322	This work
pT15-217	<i>pgpB</i> , <i>ori</i> pACYC184	W. Dowhan (18)
pLC13-12	<i>uvrC</i> , <i>ori</i> ColE1	W. Dowhan (8)
pTX277	<i>pdxB</i> , <i>ori</i> pBR322	29
pLAFRx	derivative of pLAFR, <i>ori</i> RK2	This work (14)
pGE2719	<i>metE</i> complementing clone from <i>Bam</i> HI library in pLAFRx	This work
pGE2769	<i>metE</i> complementing clone from <i>Pst</i> II library in pLAFRx	This work

$\mu$ g/ml; and chloramphenicol (CAM), 25  $\mu$ g/ml. The final concentration of isopropyl-thiogalactoside (IPTG) was 0.5 mM.

**Auxotrophic transposon insertion mutants.** GE2515 is an MG1655 derivative that has been transformed with plasmids pMS421 (Spc<sup>r</sup>) and pJD12 (Amp<sup>r</sup>) (Fig. 1). Plasmid pMS421

is a derivative of pSC101 carrying the *lacI<sup>q</sup>* gene (overproduces Lac repressor) (15). Plasmid pJD12 was derived from pMM12, a derivative of plasmid pZT344 (13). Plasmid pZT344 is a pBR322 replicon carrying a defective Tn10 element containing the *cat* gene (chloramphenicol resistance) flanked by 66 bp from the ends of Tn10. Plasmid pZT344 also contains the Tn10 transposase under the control of the *tac* promoter. To make plasmid pMM12, *Not*I and *Sfi*I restriction sites were inserted inside one 66-bp end of the Tn10dCam element, creating transposon Tn10dCamNS (3). The sequence of the polylinker in this transposon is GGATCCGGCCCCGGGGGGCCGCGGCCGC-*cat*. By using oligonucleotides, *Xba*I, *Nhe*I, and *Dra*I restriction sites were inserted between the *Not*I and *Sfi*I sites of the Tn10dCamNS element of plasmid pMM12, regenerating the *Not*I and *Sfi*I sites and creating Tn10dCamMCS. The sequence of the polylinker in this transposon is GGATCCGGCCCCGGGGGGCCTCTAGATTAAAGCTAGCGCGGC CGC-*cat*. The resultant plasmid pJD12 was used to transform MG1655 containing plasmid pMS421. Plasmid pMS421 provides the *lacI<sup>q</sup>* repressor to control expression of the transposase and subsequent activation of the transposon. Strain GE1669 is identical to GE2515 except that it contains plasmid pMM12 in place of plasmid pJD12. GE1669 was used to generate some of the transposon insertion mutants as described below.

GE2515 (or GE1669) was grown from an isolated colony overnight in L broth with CAM and spectinomycin. Then, 0.1 ml of the overnight growth was added to 5 ml of L broth-spectinomycin-CAM-IPTG and allowed to incubate at 37°C with aeration for 4 h. During this time transposition of the Tn10 derivative present on the multicopy plasmid occurred. A 0.2-ml sample of a 10<sup>-6</sup> dilution was spread on each of several L-CAM plates and incubated at 37°C overnight. At this stage, the cells contain multiple insertions in their chromosomes. The colonies were replica plated onto M63-CAM plates and placed again at 37°C overnight. Mutants with a nutritional requirement fail to grow on the M63-CAM plates. Auxonography was performed as described by Davis, Botstein, and Roth (11) to determine the

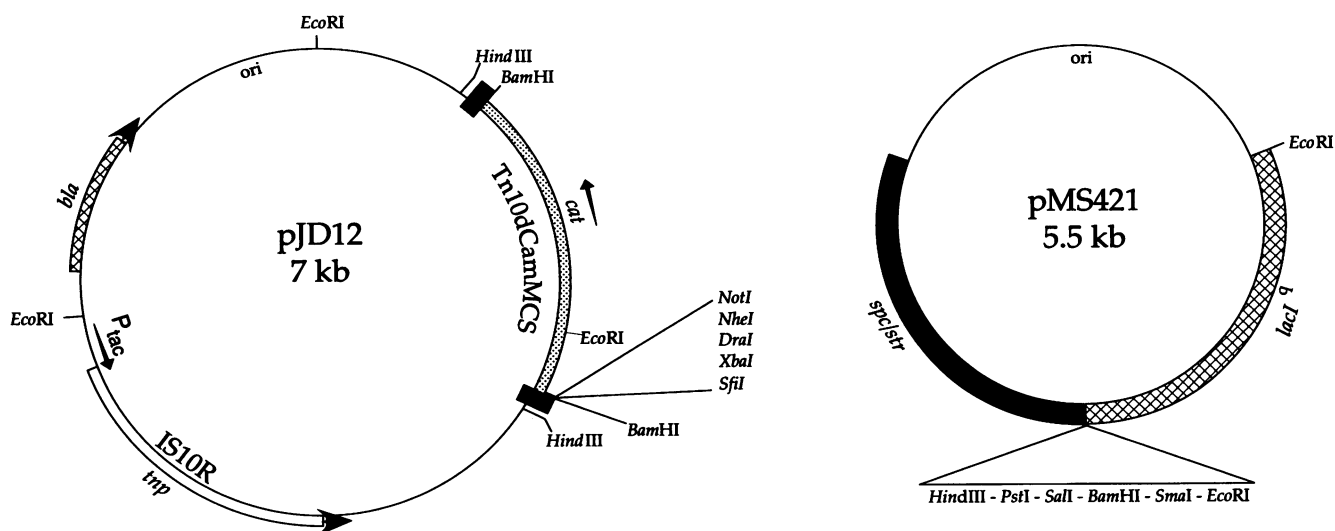


FIG. 1. Tn10dCamMCS transposition system. Plasmid pJD12 is a derivative of pZT344 (13) which has been engineered to contain rare restriction sites in the Tn10dCam element. The *tnp* gene coding for the Tn10 transposase is located outside the Tn10dCam element and is under the control of the *P<sub>lac</sub>* promoter. Plasmid pMS421 provides the *lacI<sup>q</sup>* repressor to regulate the expression of the transposase.

nutritional requirement of each mutant. Insertions that resulted in an auxotrophic phenotype were separated from other insertions in the strain by P1 transduction of MG1655 to *Cam<sup>r</sup>* and auxotrophy using a P1 lysate from the auxotrophic strain as the donor. Transductants were tested for the absence of plasmids pMS421 and pJD12 by screening for spectinomycin and ampicillin sensitivity. The presence of a single insertion was confirmed by transducing to prototrophy and screening for *CAM* sensitivity. Auxotrophs generated in this way introduce a new *NorI* site, along with other sites, at the location of the *Tn10dCamMCS* insertion. These insertions were mapped by pulsed-field gel electrophoresis (PFGE), which verified that only a single insertion was present.

**PFGE.** To prepare intact DNA in agarose plugs, cells were grown overnight in L broth, pelleted, and resuspended in 1/2 volume of PIV buffer (1 M NaCl, 10 mM Tris-HCl [pH 7.6]). Plugs were prepared by mixing 1 ml of the cell suspension with an equal volume of molten 1.6% InCert (FMC) agarose in water. Typically the cell density at this point (optical density at 600 nm) was 6. The cells in 0.8% agarose were then pipetted into plug molds (BioRad) and allowed to solidify. These plug molds make a plug that is ~250  $\mu$ l in volume and is 20 by 9 by 1 mm in size. The agarose plugs were incubated at 37°C with gentle shaking overnight in EC lysis solution (6 mM Tris-HCl [pH 7.6], 1 M NaCl, 100 mM EDTA [pH 7.5], 0.5% Brij 58, 0.2% deoxycholate, 0.5% Sarkosyl, 1 mg of lysozyme per ml, 20  $\mu$ g of RNase per ml) using about 5 ml per two plugs. The EC lysis solution was replaced with ESP (0.5 M EDTA [pH 9 to 9.5], 1% Sarkosyl, 50  $\mu$ g of proteinase K per ml), and incubation was continued at 50°C with gentle shaking overnight. The plugs were washed three times for 30 min each in at least 10 volumes of TE (10 mM Tris-HCl [pH 7.5], 1 mM EDTA). At this point, the plugs may be stored at 4°C. By using gloves, a plug was placed on a microscope slide and cut with a glass coverslip. A small slice of a plug was placed in 250  $\mu$ l of the appropriate enzyme buffer, and enzyme was added. Ten to twenty units of *NorI* was sufficient for complete digestion in 1 h at 37°C. After complete digestion, the buffer was removed and the plug was incubated in TE for 30 min. The plug was loaded into an agarose gel by either inserting the solid plug in a well or melting the plug and gently pipetting it in. Generally, 40 ng of DNA was sufficient for viewing all the fragments of the digest.

Two types of agarose were used for contour-clamped homogeneous electric field (CHEF) gels. SeaPlaque GTG (FMC) was used for measurements of DNA fragment sizes because bands were sharper in this matrix. It was also used when bands were to be recovered from gels (e.g., to be used as probes), because this brand is low-melting-point agarose. FastLane agarose (FMC) was used when DNA was to be transferred to membranes for Southern blotting, because it gave more efficient transfer. Because of the difference in the gel matrix, FastLane gels were run at a slightly faster pulse time and for a shorter duration than SeaPlaque gels.

Samples were subjected to PFGE using the CHEF system (7). Typical running conditions in the DRII apparatus from Bio-Rad with SeaPlaque agarose were 200 V and 0.08 A (constant voltage) in 0.5 $\times$  TBE (1 $\times$  TBE is 0.9 M Tris-HCl [pH 7.6]–0.9 M boric acid–1.0 mM EDTA) at 8°C with recirculating buffer for about 40 h. When FastLane agarose was used, the buffer was 0.25 $\times$  TBE and the run length was 20 h. Most gels were run with ramping, in which the pulse time was continuously increased throughout the run. A pulse time of 5 to 40 s was standard with SeaPlaque agarose and

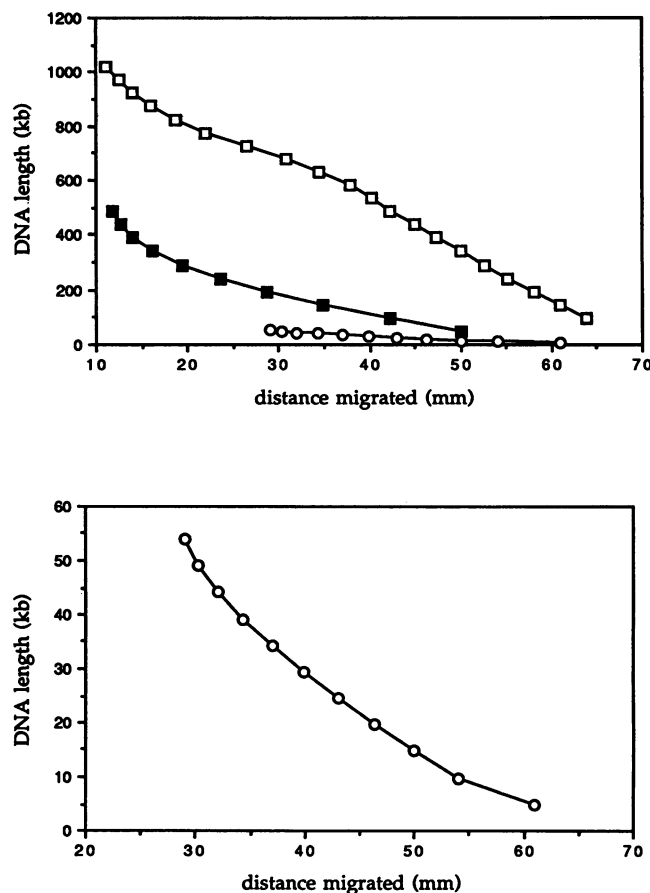


FIG. 2. Separation of DNA size standards by PFGE. The separation of lambda concatemer DNA size standards under various pulse times is shown. The gel for each was 1.2% SeaPlaque GTG agarose in 0.5 $\times$  TBE. The gels were run for 40 h at 200 V and a temperature of 8°C with recirculating buffer. Pulse times in each case were linearly ramped beginning with the shortest pulse time and ending with the longest. Symbols:  $\square$ , 75- to 100-s pulse;  $\blacksquare$ , 5- to 40-s pulse;  $\circ$ , 0.5- to 5-s pulse. The lower graph shows the short pulse time sample with an expanded scale.

gave good resolution of fragments from about 50 to 300 kb. A pulse time of 75 to 100 s was adequate for resolving fragments above 300 kb. The best separation of fragments smaller than 50 kb was obtained when a pulse time of 10 to 20 s for 10 h was followed by a pulse time of 1 to 10 s for an additional 20 h. Figure 2 illustrates the resolution obtained under these conditions of CHEF gel electrophoresis. Greater resolution of specific size ranges may be attained by limiting the pulse times to a smaller range or by running the gel at one continuous pulse time for a period of time such that the fragment range of interest can migrate further through the gel. The best resolution of a wide range of fragments, from 10 kb up to approximately 400 kb, on a single gel was attained by modifying the CHEF DRII apparatus to allow large-format gels. Gels of 24 by 20.5 cm were accommodated by drilling two 9-mm-deep holes with a 5/32-in. (ca. 0.4-cm) drill bit into the Plexiglas body of the unit for the gel support posts. A large-format 1.2% SeaPlaque GTG agarose gel run with standard conditions and a pulse time of 5 to 40 s for 60 h resulted in straight lanes with good resolution of fragments ranging from 9 kb up to about 400 kb.

**Determination of DNA fragment sizes.** After electrophoresis, gels were stained with ethidium bromide and photographed on a UV transilluminator. By using the photograph, the distance between the leading edge of the sample well and the leading edge of the bands was measured to the nearest 0.1 mm. Preliminary experiments with *NotI* digests of MG1655 DNA showed that leading-edge measurements on CHEF gels were less affected by overloading than either lagging-edge or center measurements. The size of unknown bands was determined by interpolation (see below) using DNA size standards in the same gel. For most gels,  $\lambda$  DNA concatemers (unit size = 48.5 kb) obtained from FMC were used as size standards. For size determination of fragments less than 50 kb in length, a series of size standards between 10.1 and 48.5 kb was constructed by cleaving  $\lambda$  DNA with the following restriction endonucleases: *ApaI*, *KpnI*, *NarI*, *SacI*, and *Sall*. These were combined to produce a size standard ladder with fragments of the following sizes: 10.0, 15.3, 17.1, 22.6, 24.8, 29.9, 38.4, 45.7, and 48.5 kb. The last fragment is uncut  $\lambda$  DNA.

Interpolation of fragment sizes was based on the observation that the relationship between band mobility and size is approximately linear for the size standards used in this study with appropriate pulse times and ramping (see discussion of Fig. 2 above). Given this relationship, the size of an unknown fragment was determined from the two flanking size standard fragments by using a two-point linear model. Several gels were also analyzed by fitting the size standards to a computer-generated polynomial equation and using this to predict sizes of unknown fragments. Sizes determined in this way were not significantly different from those determined by two-point linear interpolation. Gel mobility data were not consistent with a logarithmic or semilogarithmic model.

**Hybridization of *NotI* fragment probes to *E. coli* gene mapping membranes.** *NotI* fragments were mapped with respect to the genomic restriction map of Kohara et al. (19, 20) by hybridizing  $^{32}\text{P}$ -labelled fragment probes to the Gene Mapping Membrane from Takara Biomedicals. The Gene Mapping Membrane contains DNA from each of 476 phage  $\lambda$  clones (the miniset of the Kohara et al. library) arranged in a grid. A radiolabelled genomic probe will hybridize to a series of the clones, and thus its location on the genomic restriction map can be determined.

To prepare probes, *NotI* fragments from SeaPlaque agarose gels were excised and radiolabelled with [ $^{32}\text{P}$ ]dCTP by using the random hexanucleotide priming method (Multiprime DNA labelling system; Amersham). Probes were hybridized to the filters in Hoefer Scientific Instruments PR800 Hybrid-Ease chambers. Following incubation for at least 2 h at 65°C in prehybridization solution (1 M NaCl, 50 mM Tris-HCl [pH 8], 1% sodium dodecyl sulfate [SDS], 200  $\mu\text{g}$  of heat-denatured sheared salmon testis DNA per ml), filters were incubated overnight at 65°C in 20 ml of hybridization solution (1 M NaCl, 50 mM Tris-HCl [pH 8], 1% SDS, 10% dextran sulfate, 100  $\mu\text{g}$  of heat-denatured sheared salmon testis DNA per ml) with approximately  $10^6$  dpm of heat-denatured probe added per ml of hybridization solution. Filters were washed (15 min per wash) once with  $2\times$  SSC–0.5% SDS at 65°C, twice with  $1\times$  SSC–0.5% SDS at 65°C, and twice with  $0.1\times$  SSC–0.1% SDS at room temperature ( $1\times$  SSC is 0.15 M NaCl plus 0.015 M sodium citrate). They were then exposed to X-ray film at –80°C, usually overnight. After autoradiography, the filters were stripped in boiling 0.5% SDS and then rehybridized.

**Southern blots of CHEF gels.**  $\lambda$  concatemers and *NotI* digests of MG1655 were separated by CHEF gel electropho-

resis in 1% FastLane agarose (FMC)–0.25 $\times$  TBE. A standard pulse time of 5 to 35 s with ramping was used, and the gels were run for 20 h at 200 V. Separated fragments were transferred to a GeneScreen Plus Hybridization Transfer Membrane (NEN-DuPont) by using the protocol outlined in reference 4. Gels were stained with 1  $\mu\text{g}$  of ethidium bromide per ml for 30 min with constant shaking. The gel was irradiated for 60 s with a 254-nm light source at an intensity of 1.4 mW/cm<sup>2</sup>. After soaking in 0.4 N NaOH for 15 min, the DNA was transferred onto the membrane by using 2 liters of 0.4 N NaOH as solvent by standard capillary transfer methods. Following transfer for at least 24 h, the membrane was neutralized in 0.5 M Tris (pH 7). The procedures outlined in the previous section were used for labelling of probe, hybridization, autoradiography, and stripping of the Southern blots.

**Computer analysis of DNA sequence data bases and assignment of physical map coordinates.** DNA sequences were analyzed with the Genetics Computer Group Sequence Analysis Software Package, version 7.0 (12), running on a DEC MicroVax 3900. The GCG Find program was used to locate *NotI* cleavage sites in *E. coli* chromosomal sequences from two data bases: GenEMBL and EcoSeq2 (23). The GenEMBL DNA sequence data base is provided with the GCG package and is composed of sequences from release 67.0 of GenBank, supplemented with unique sequences from release 26.0 of the EMBL data base. EcoSeq2 is an *E. coli* chromosomal DNA sequence data base consisting of all published sequences in GenBank and EMBL (up to 1 August 1991), as well as unpublished sequences. Contiguous sequences have been joined to yield larger melds, so that all sequences are represented only once in the data base. This data base contains approximately 35% of the sequence of the *E. coli* chromosome. We obtained EcoSeq2, along with Ecoli4.map (described below) and several documentation files, by anonymous Ftp from the National Center for Biotechnology Information Internet computer ncbi.nlm.nih.gov, as detailed by Rudd et al. (23). We also downloaded and compiled the C source code for four utility programs that accompany EcoSeq2 and Ecoli4.map: PrintMap, ProbeMaker, MapSearch, and AlterMap.

Physical map coordinates for *NotI* sites and transposon insertions were assigned on the basis of the coordinate system of Ecoli4.map (23). Ecoli4.map is a revised version of the genomic restriction map of Kohara et al. (19, 20). Following the conventions of the original Kohara map, restriction sites of the eight restriction enzymes have been assigned coordinates, with the beginning of the *thr* operon as the 0-kb reference point and proceeding in the direction of transcription of the *thr* operon. Consequently, 0 kb on the physical map is equivalent to 0 min on the genetic map, and the maps run in the same direction. Ecoli4.map was constructed by first aligning the sequences of EcoSeq2 with the Kohara map. Then, the coordinates of the restriction sites were altered to make them consistent with the sequence. Approximately 31% of the resulting Ecoli4.map is based on sequence data, with the remainder being based on the Kohara restriction map. Another 4% of the chromosome is sequenced but could not be unambiguously placed on the map. The estimated size of the *E. coli* chromosome, based on Ecoli4.map, is 4,671.0 kb. One important feature of Ecoli4.map that differentiates it from the original Kohara map is that the chromosomal segment between *rrnD* and *rrnE*, which is inverted in strain W3110 and in the original Kohara map, has been reinverted to the wild-type order in Ecoli4.map. A second modification is the removal in

Ecoli4.map of four of the five repeats of the *tdc-rnpB* region, represented in five copies in the Kohara map (23, 24). Thus, Ecoli4.map corresponds more closely to the chromosome of strain MG1655.

The first step in assigning coordinates to *NotI* sites and transposon insertions was to assign coordinates to sequenced *NotI* sites present on Ecoli4.map. Exact coordinates for *NotI* sites were determined by first locating the sequences containing *NotI* sites on Ecoli4.map and then taking into account the location of the site within the sequence. Next, coordinates were assigned to transposon insertions in sequenced genes on Ecoli4.map. While exact coordinates exist for these genes or operons, the precise location of the insertion in a gene is not known. The center of the gene or operon was used as the coordinate for the insertion. Since some of these genes or operons are several kilobases in size, this limits the precision of transposon coordinate assignments.

Transposon insertions in genes that have not been placed in Ecoli4.map were assigned coordinates next, on the basis of the position of the affected genes on the genetic map. Because of the imperfect correlation between the genetic and physical maps of *E. coli*, these coordinates are less exact than those for placed genes. In one case, that of the *metE::Tn10dCamMCS* strain, one of the *NotI* fragments adjoining the insertion was used to probe the Gene Mapping Membranes. Therefore, coordinates for this insertion could be assigned with more certainty.

Finally, two methods were used to assign approximate coordinates to *NotI* sites that lie in regions that are not placed in Ecoli4.map. If a *NotI* fragment had been used to probe a Gene Mapping Membrane, coordinates were assigned for the endpoints of the fragment on the basis of these results. Because of the close relationship between Ecoli4.map and the Kohara phage library, these coordinates could usually be assigned within a range of 2 to 6 kb. For the remainder of the unsequenced sites, coordinates were assigned on the basis of nearby sequenced sites or transposon insertions, combined with data from PFGE analysis. This method of assigning coordinates is less exact than that for sequenced sites or sites placed with Kohara filter hybridization. Therefore, the range for these sites is larger, from 8 to 18 kb.

## RESULTS AND DISCUSSION

***NotI* cleavage map of strain MG1655.** Four approaches were combined to place *NotI* cleavage sites on the genetic and physical maps. (i) Transposon insertions were isolated that caused auxotrophic mutations. This phenotype allowed each insertion to be located on the genetic map by virtue of its particular nutritional requirement, once its location had been narrowed to a particular region by restriction mapping. The transposon insertion also introduced a new *NotI* site, resulting in the loss of one wild-type *NotI* fragment and the appearance of two new fragments. The distance from the insertion to the *NotI* sites at the ends of the wild-type fragment could be calculated from the sizes of the new fragments. Since the location of the insertion on the genetic map was known, this allowed a correlation between *NotI* sites and the genetic map to be derived. (ii) *NotI* restriction fragments were isolated following PFGE and used as probes of filters containing the ordered miniset of phage  $\lambda$  clones (21). Typically a fragment would hybridize with a contiguous set of phage clones, allowing the endpoints of the fragment, and thus the *NotI* sites, to be mapped. Occasionally the gel

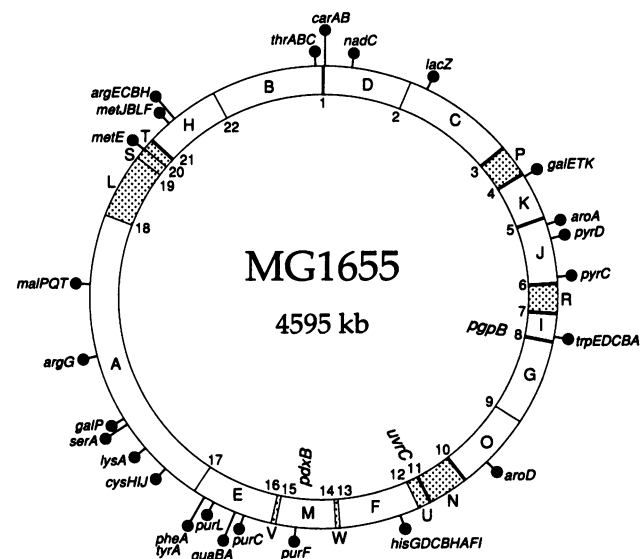


FIG. 3. *NotI* restriction map of *E. coli* MG1655. The map is oriented so that 0 min on the genetic map is located at the top of the map at locus *thrABC*. *NotI* fragments are assigned letters for consistency with the EMG2 map of Smith and coworkers (10, 27). *NotI* sites are also numbered on the inside of the map in a clockwise orientation from 0 min. Transposon insertion locations are designated by lollipops and the inactivated genetic loci. *NotI* sites are represented by thick lines for sequenced sites or thin lines for unsequenced sites. Fragments that have been placed by hybridization to Gene Mapping Membranes are shaded. Clones that were used to place some of the fragments are indicated on the inside circumference of the map.

band was a doublet. In this case, two separate groups of phage clones would hybridize and both fragments could be mapped. (iii) Recognition sites for *NotI* were identified in the collection of *E. coli* sequences assembled by Rudd et al. (22, 23). This immediately placed sites on the genetic map and the associated physical map derived from these sequences and the phage  $\lambda$  miniset by Rudd et al. (22, 23). (iv) Cloned genes, whose location on the genetic map was known, were used as probes in Southern blots of pulsed-field gels of *NotI* digests. In some cases the cloned sequences contained *NotI* sites (linking clones) and hybridized to two fragments, establishing that these fragments were contiguous.

The map derived from these approaches is shown in Fig. 3. The map is 4,595 kb in length. We have labelled fragments to correspond to the map of EMG2 derived by Smith and coworkers (10, 27). They labeled fragments on the basis of their sizes, starting with the largest fragment. Because of strain differences, this relation does not hold for MG1655, and our fragment names were chosen only to avoid confusion with the earlier map. The correlation between the sets of fragments in these strains is shown in Table 2.

The location of *NotI* fragments and sites is based on different approaches, as described above, and the evidence for the location of each fragment and site is summarized in Tables 3 and 4. In all, there are 22 *NotI* fragments and 18 of the fragments are placed by multiple pieces of evidence. A total of 26 transposon insertions were used in the mapping, and these occurred in 13 different fragments. Nine fragments were used as probes of the ordered  $\lambda$  library. Of the four fragments whose location is mapped by a single piece of

TABLE 2. *NotI* fragment size and order of fragments in MG1655 and EMG2

MG1655		EMG2 <sup>a</sup>	
Fragment	Size (kb)	Fragment	Size (kb)
A	992	A	1,000
B	358	B	360
C	358	C	360
D	279	D	275
E	263	E	250
F	251	F	250
G	273	G	245
H	245	H	240
I	90	I	230
J	214	J	210
K	153	K	205 <sup>b</sup>
L	207	L	203
M	190	M	191
N	129	N	130
O	251	O	106
P	103	P	100
Q		Q	100 <sup>c</sup>
R	96	R	95
S	36	S	43
T	38	T	43
U	41	U	40
V	14	V	20
W	14		
Total size	4,595	Total size <sup>d</sup>	4,547

<sup>a</sup> The EMG2 fragments are based on the work of Condemine and Smith (10).

<sup>b</sup> Fragment K of EMG2 contains a 48.5-kb  $\lambda$  prophage that is absent in MG1655.

<sup>c</sup> Fragment Q in EMG2 is the F factor that is absent in MG1655 and not included in the total.

<sup>d</sup> The total size does not include the  $\lambda$  prophage or the F factor of EMG2.

data, three (L, V, and W) have been mapped by hybridization to filters with the phage  $\lambda$  library, and thus their endpoints are well localized. The remaining fragment (F) is located between two fragments that have been mapped to the  $\lambda$  library (fragments U and W), and thus its location is likewise well established.

One anomaly was noted when using the DNA sequence information to aid in the derivation of the map. There is a *NotI* recognition site in the DNA sequence of the *bioA* gene, located at 17.5 min on the genetic map, that is not present in strain MG1655. This difference could be due to sequencing error but may also represent a bona fide difference between MG1655 and the K-12 strain used to sequence this region. Otherwise there was a very good correlation between sites predicted from the DNA sequence and what was observed. In general, distances estimated between sites from the DNA sequence coordinate system were found to be slightly larger than the distances measured by PFGE. This may be due to the fact that only about one-third of the chromosome is sequenced, and the distances between sites in the remaining two-thirds can only be estimated.

The EcoSeq2 sequence collection of Rudd et al. has been correlated with the  $\lambda$  clones of Kohara et al. and this forms the basis for a coordinate system for the *E. coli* physical map (see Materials and Methods). We have used this system to assign coordinates to all of the *NotI* cleavage sites (Table 4) and transposon insertions (Table 5) as described in Materials and Methods. Table 4 also shows phage clones and DNA sequences that carry *NotI* sites.

TABLE 3. Summary of evidence for the *NotI* map of MG1655

Fragment	Transposon insertion(s)	Sequenced endpoint(s) <sup>a</sup>	Probe(s) <sup>b</sup>
A	<i>cysHIJ</i> , <i>lysA</i> , <i>serA</i> , <i>galP</i> , <i>argG</i> , <i>malPQT</i>		
B	<i>thrABC</i>	d	
C	<i>lacZ</i>	d	
D	<i>carAB</i> , <i>nadC</i>	u	
E	<i>purC</i> , <i>guaBA</i> , <i>purL</i> , <i>tyrA</i> , <i>pheA</i>		
F	<i>hisABCDEFGHI</i>		
G		u	<i>pgpB</i>
H	<i>metJBLF</i> , <i>argECBH</i>	u	
I	<i>trpEDCBA</i>	u,d	<i>pgpB</i>
J	<i>aroA</i> , <i>pyrD</i> , <i>pyrC</i>	u,d	
K	<i>galETK</i>	u,d	
L			Frag. <i>pdxB</i>
M	<i>purF</i>		
N		u,d	Frag., <i>uvrC</i>
O	<i>aroD</i>	d	<i>uvrC</i>
P		u,d	Frag.
R		u,d	Frag.
S	<i>metE</i>		Frag.
T		d	Frag.
U		u	Frag.
V			Frag.
W			Frag.

<sup>a</sup> d; downstream site is sequenced; u; upstream site is sequenced.

<sup>b</sup> *NotI* fragments of the MG1655 chromosome were used to probe Gene Mapping Membranes (Frag.), or plasmid clones of chromosomal genes were used to probe Southern blots of MG1655 digested with *NotI*.

**Comparison of strains MG1655 and EMG2.** Out of all extant strains of *E. coli* K-12, strain EMG2 (9) is considered to most closely resemble the original wild-type *E. coli* K-12 isolate (1). It is lysogenic for phage  $\lambda$  and carries the F factor as a plasmid. Strain MG1655 is two steps removed from wild type, having been cured of  $\lambda$  in the first step and F in the second (2). A gel showing the *NotI* restriction patterns of these strains is shown in Figure 4. The  $\lambda$  prophage in EMG2 is in fragment K (205 kb), which is missing in MG1655 and is replaced by a 153-kb fragment K. The F factor in EMG2 is fragment Q, which appears as part of a doublet at around 100 kb. This doublet is replaced by a single band at this position in MG1655. A number of other differences can also be seen between these two strains. In Fig. 5 the map of MG1655 derived here is aligned with the EMG2 map (10, 27). Most fragments are similar in the two strains, although there are several examples of small size differences (a few kilobases). Such variation can be rationalized on the basis of movement of insertion sequences for example. However, one region that shows a major difference in restriction pattern is the segment from approximately 25 to 40 min (fragments I, G, and O). No simple rearrangement, such as an inversion, was consistent with the patterns that were observed, and we were unable to devise a simple hypothesis to account for the difference between strains. It may be significant that there are several cryptic prophages in this area (5). A second major difference between the strains is seen near 50 min on the map, where MG1655 has an additional fragment (fragment W). This difference could be caused by a point mutation creating a new restriction site. A final quantitative difference is in the size of the smallest fragments (i.e., less than 50 kb) reported for the two strains (Table 2). Since in our experiments the fragments from the two strains electrophoresed with the same mobility (data not shown), the

TABLE 4. Physical and genetic map locations of *NotI* cleavage sites

Feature	Location on physical map (kb) <sup>a</sup>	Location in Kohara phage miniset <sup>b</sup>	Location on genetic map (min)	Gene	GenBank sequence and position in sequence <sup>c</sup>	Evidence <sup>d</sup>
sN1	25.087	<b>[103]22B12</b>	0.4	<i>ileS</i>	ECOLSP 381	S,t
sN2	312 ± 8	<b>[131]21C10</b> [132]5E5 [133]3E3 [134]9G4 <b>[169]15D7</b>	7.0 ± 0.3			T,A
sN3	677.816		15.0	Orf2 upstream of <i>mrdA</i> ( <i>pbp A</i> )	ECOPBPA 720	S,t,a
sN4	787.394	<b>[177]18F11</b> <b>[178]10G5</b>	16.8	<i>tolA</i>	ECOTOLAB 946	S,t,a
sN5	941.827	<b>[213]1H1</b> <b>[214]1F10</b> <b>[215]E6H3</b> <b>[234]E3G11</b> <b>[235]14C1</b>	19.8 ± 0.1	<i>lrp</i>	lrp-eco 1147 <sup>e</sup>	S,t,a
sN6	1,157.164	<b>[245]7C10</b> <b>[253]4F1</b> <b>[254]13F9</b> <b>[255]18B6</b> <b>[256]E14F6</b>	24.8	<i>rpmF</i>	ECORPMFA 1016	S,t,a,g
sN7	1,252.025		26.3	<i>treA</i>	ECOTREA 1702	S,t,a,g
sN8	1,350 ± 8 <sup>f</sup>	<b>[303]5F9</b> <b>[304]1F9</b> <b>[305]2H2</b> <b>[330]3E12</b> <b>[331]9F2</b> <b>[341]16B12</b>	28.5	<i>pgpB</i>	M23628 517	T,A,s
sN9	1,620 ± 5		34.0 ± 0.5			T,A
sN10	1,875.420 or 1,874.576 <sup>g</sup>		39.3	<i>gap</i>	ECOGAP 1070	S,T,A,G
sN11	2,011.382		42.1	Orf1 upstream of <i>uvrC</i>	ECOUVRC 1160	S,g,a
sN12	2,050 ± 3	<b>[344]25D8</b>	43.1			G,t,a
sN13	2,299 ± 3	<b>[371]20F6</b>	47.3 ± 0.2			G,t,a
sN14	2,318 ± 3	<b>[372]19D1</b> <b>[373]22A6</b> <b>[413]7G9</b> <b>[414]10D3</b> <b>[415]12H4</b>	47.5 ± 0.2			G,t,a
sN15	2,513 ± 2		51.3 ± 0.2			G,t,a
sN16	2,523 ± 2	<b>[440]24G1</b> <b>[441]4C9</b>	51.5 ± 0.2			G,t,a
sN17	2,782 ± 8	<b>[575]5D2</b> <b>[554]10F4</b> <b>[551]1C10</b> <b>[550]16G1</b> <b>[547]8D12</b> <b>[546]10F1</b>	57.0 ± 0.2			T,A
sN18	3,801 ± 5		80.9 ± 0.3			G,t,a
sN19	4,005 ± 1		85.4 ± 0.1			G,t,a
sN20	4,046 ± 1		86.6 ± 0.2			C,G,a
sN21	4,082.439	<b>[640]1A11</b> <b>[641]12A6</b> <b>[642]10H5</b> <b>[643]12H2</b>	87.6	<i>glnL</i>	ECOGLN 2634	S,g,t,a
sN22	4,333 ± 6		92.7 ± 0.3			T,A

<sup>a</sup> Coordinates based on Ecoli4.map. 0 kb is defined as the beginning of the *thr* operon (0 kb = 0 min).

<sup>b</sup> Phages in bold type contain *NotI* cleavage sites by direct evidence (i.e., the site has been sequenced or an adjacent *NotI* fragment has been used to probe a Gene Mapping Membrane). Phages in small type potentially contain *NotI* cleavage sites as determined by indirect evidence (i.e., site placement by PFGE data).

<sup>c</sup> Positions refer to the beginning of the *NotI* cleavage site with respect to the beginning of the GenBank (or EcoSeq2) sequence.

<sup>d</sup> Evidence used to assign coordinates to the sites. Uppercase letters represent evidence used to assign the coordinate or coordinate range to the site. Lowercase letters represent evidence that supports this coordinate or range. A or a; location of site deduced from the location of adjacent sites whose location was directly determined; C or c; location of site determined by restriction mapping of a cosmid clone of a region containing a *NotI* site; G or g; location of site determined by hybridization of an MG1655 *NotI* fragment to a Gene Mapping Membrane; S or s; site is present in a published DNA sequence which has been placed on Ecoli4.map, or a placed sequence without any *NotI* sites restricts the location of a site to outside of the sequenced area; T or t; location of site deduced by analyzing *NotI* digests of strains containing mapped transposon insertions.

<sup>e</sup> The sequence lrp-eco is an unpublished sequence that is part of EcoSeq2.

<sup>f</sup> The sequence M23628 containing the *pgpB* gene is part of EcoSeq2 but is not placed on Ecoli4.map. Therefore, physical map coordinates can not be accurately assigned to this sequence.

<sup>g</sup> The sequence ECOGAP can be placed on Ecoli4.map based on the pattern of recognition sites for the eight enzymes used by Kohara et al. (20) to construct their *E. coli* K-12 physical map, combined with our *NotI* mapping data, but its orientation cannot be determined. Therefore, there are two possible coordinates for the *NotI* site contained in this sequence.

TABLE 5. Physical and genetic map locations of transposon insertions

Gene(s)	Location on physical map (kb) <sup>a</sup>	Location on genetic map (min)	Location on NotI map (fragment)
<i>thrABC</i>	2.5 ± 2.5	0.0	B
<i>carAB</i>	31.8 ± 2.3	0.7	D
<i>nadC</i>	115.0 ± 3.4	2.7	D
<i>lacZ</i>	371.1 ± 1.6	8.0	C
<i>galETK</i>	798.9 ± 2.1	17.0	K
<i>aroA</i>	967.7 ± 0.7	20.2	J
<i>pyrD</i>	1,012.4 ± 0.5	21.2	J
<i>pyrC</i>	1,131.3 ± 0.6	23.7	J
<i>trpEDCBA</i>	1,327.5 ± 3.4	28.0	I
<i>aroD</i>	1,785.4 ± 0.5	37.2	O
<i>hisGDCBHAFI</i>	2,107.0 ± 3.7	44.0	F
<i>purF</i>	2,448.7 ± 0.4	50.0	M
<i>purC</i>	2,602.8 ± 0.5	53.2	E
<i>guaBA</i>	2,637.4 ± 1.6	53.9	E
<i>purL</i>	2,700.6 ± 2.0	55.2	E
<i>pheA</i>	2,752 ± 9*	56.6	E
<i>tyrA</i>	2,752 ± 9*	56.6	E
<i>cysHIJ</i>	2,906 ± 2	59.3	A
<i>lysA</i>	2,993.8 ± 0.7	61.3	A
<i>serA</i>	3,073.1 ± 0.6	63.0	A
<i>galP</i>	3,111 ± 7*	63.8	A
<i>argG</i>	3,336.2 ± 0.7	69.0	A
<i>malPQT</i>	3,571.1 ± 3.9	75.3	A
<i>metE</i>	4,039 ± 2†	86.3	S
<i>metJBLF</i>	4,157.6 ± 2.8	89.0	H
<i>argECBH</i>	4,182.7 ± 2.7*	89.5	H

<sup>a</sup> Coordinates based on Ecoli4.map. 0 kb is defined as the beginning of the *thr* operon (0 kb = 0 min). \*, gene has not been sequenced or placed on the physical map. We assigned putative physical map coordinates based on the location on the genetic map. †, fragment with a transposon insertion as its endpoint was used to probe the Gene Mapping Membranes. We used this information to assign coordinates to the transposon insertion.

differences in reported size most likely reflect differences in electrophoresis or measuring techniques.

**Restriction patterns of other strains.** The *E. coli* strain W1485 is the intermediate between the wild type and MG1655 (2). It has been cured of  $\lambda$  but still carries the F plasmid. It was thus of interest to compare its pattern with those of these strains. As can be seen in Fig. 4, W1485 lacks the  $\lambda$  prophage (loss of the 205-kb fragment and presence of the strain of the 153-kb fragment) but still carries the F factor (presence of the doublet at ~100 kb). However, it is also apparent that the W1485 strain used in this study is not a simple intermediate between the other strains. In particular, in the region from 240 to 280 kb and the region at 85 kb, W1485 differs from EMG2 and MG1655 in a way that is inconsistent with its being a simple intermediate. It may be that all of these strains have changed after laboratory culture.

We also investigated the structure of three commonly used strains, W3110, AB1157, and MC4100. Strain W3110 is the strain used for the construction of the  $\lambda$  library by Kohara et al. It is known to have an inversion between two rRNA operons (*rrnD* and *rrnE*) (17) as well as an amplification of a sequence at the *tdc* and *rnpB* genes and a translocation (22–24, 28) (Fig. 5). The inversion and the amplified region have been corrected in the sequence data base assembled by Rudd et al. The inversion should result in the loss of fragments A (992 kb) and H (245 kb) and the appearance of two new large fragments, seen at the top of the gel in Fig. 4. When different pulse times were used to allow more accurate

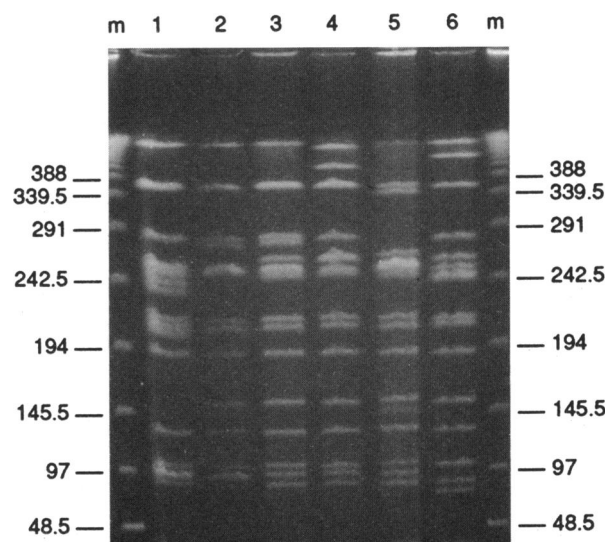


FIG. 4. CHEF gel of DNA from various *E. coli* strains digested with NotI. Photograph of a 1.2% SeaPlaque GTG agarose gel stained with ethidium bromide. NotI fragments of chromosomal DNA from various *E. coli* strains were separated by PFGE on a CHEF DR11 unit. Lanes m, lambda concatemer DNA size standards from FMC. The size of each marker in kilobases is indicated on both sides of the photograph. DNA: lane 1, EMG2; lane 2, W1485; lane 3, MG1655; lane 4, W3110; lane 5, AB1157; lane 6, MC4100.

measurement of these fragments, the fragments were observed to be 806 and 419 kb. However, taking into account the 50-kb amplification mentioned above, the map presented for MG1655 predicts the fragments to be 870 and 454 kb. This suggests there may be other differences between these strains in this region. From the gel of Fig. 4, it can also be seen that W3110 is missing fragment G (273 kb) and has an extra band slightly lower, forming a doublet at around 260 kb. Fragment G contains the Rac cryptic prophage, and thus it is possible that this prophage has been lost in W3110. Minor fragment size differences are also apparent at the 90- to 100-kb region.

Strain AB1157 has been used widely for genetic analysis, particularly in the areas of DNA recombination and repair. It is also shown in Fig. 4. This strain is known to have lost the Rac prophage, and, as predicted, it does not have fragment G of MG1655 but has acquired an extra fragment of slightly smaller size (260 kb). However, AB1157 is also missing fragment D (279 kb) of the MG1655 map. This strain has a deletion in fragment D,  $\Delta(gpt-proA)$ ; however, these markers are only about 3.6 kb apart. Thus this deletion may be considerably larger and the resulting fragment may run on top of another fragment. One other difference noted with this strain is that the two 358-kb fragments, B and C, that are not resolved in MG1655 are separated in AB1157 (Fig. 4). This suggests a small size change in one of the fragments, possibly due to another deletion or movement of an insertion sequence.

The final strain that was analyzed is MC4100, which is widely used in genetic analysis, particularly for *lac* fusions. This strain contains the deletion  $\Delta(argF-lac)U169$ , and, as shown in Fig. 5, this deletion is expected to fuse fragments C and D, resulting in a hybrid CD fragment. On the basis of the coordinates of the *argF* and *lac* genes, the CD fragment would be about 560 kb if the deletion did not extend beyond

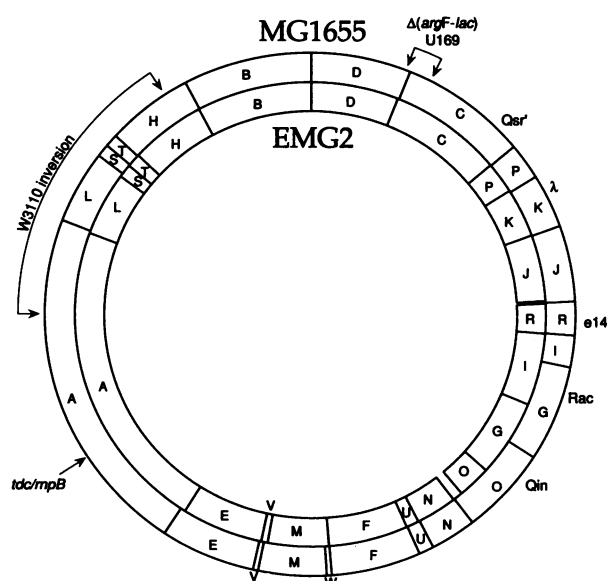


FIG. 5. Comparison of the *NotI* maps of MG1655 and EMG2. The map of MG1655 is represented by the outside ring, and that of EMG2 (10, 27) is represented by the inside ring. The fragments are drawn to scale with fragment K of EMG2 reduced by 48.5 kb to reconcile this fragment with fragment K of MG1655, which is missing prophage  $\lambda$ . The maps are very similar with the exception of the region around 25 to 35 min which includes fragments I, G and O. MG1655 also contains one fragment of 14 kb (W) which is absent in EMG2. EMG2 also contains the F factor, which has been cured in MG1655 and is not shown on this map. The positions of the excisable cryptic element e14 and the prophages Rac, Qsr', and Qin are also indicated on the map. The positions of the inversion between *rrnE* and *rrnD* in W3110 is indicated, as well as the position of the *tdc-rnpB* region at 3,281.4 kb (23). In the W3110 strain mapped by Kohara et al., this region of about 14 kb is repeated five times. Three of the four additional copies are located between 3,144.5 and 3,184.8 kb, and the fourth is located between 3,318.0 and 3,329.0 kb (23). The  $\Delta(\text{argF-lac})\text{U169}$  deletion of strain MC4100 is also shown on the map.

these genes. The CD fragment is seen in Fig. 4 and was found to be 532 kb (by using a different pulse time), suggesting the deletion is considerably larger. Two other differences were seen with MC4100 as compared with MG1655. First, the doublet at 251 kb (fragments F and O) was reduced to a singlet and a new band, forming a doublet with fragment H at 245 kb, was evident. It should be noted that fragment O has a cryptic prophage (Qin) but whether the change is due to the loss of this prophage or to another change in F or O is not known. Finally, fragment I (90 kb) was missing and a new fragment of 78 kb had appeared.

These results demonstrate that the map of MG1655, as well as that of EMG2, can be applied to other commonly used *E. coli* strains. In most regions of the chromosome there is consistency between the different strains. The region from 240 to 280 kb is one exception in which different patterns of bands exist. The other region, from 90 to 106 kb, appears to be consistent between most strains. However, the three bands in this region from EMG2 (O-P-R) and MG1655 (P-R-I) do not exactly correspond in gel order or chromosome location. Future experiments will determine whether the bands in other strains correspond to those from EMG2 or MG1655.

There are several potential applications of a restriction

enzyme cleavage map such as the one presented here. Mutations can be rapidly and accurately mapped with the use of the Tn10dCamMCS system. This is true both for transposon-generated mutations (as demonstrated by the series of auxotrophic mutations described above), and for mutations which are linked to a transposon insertion. Analysis of chromosomal rearrangements such as insertions, deletions, duplications, and inversions is also facilitated by such a map. Several of these rearrangements were noted (as described above) when several common *E. coli* laboratory strains were compared. As other restriction enzyme cleavage maps are constructed, the utility of these maps in genetic and physical analysis of the *E. coli* K-12 genome can be further increased.

#### ACKNOWLEDGMENTS

B. Bachmann's informative explanation of the history and pedigrees of several common K-12 strains was most helpful. We thank A. Noda and J. Lupski for the use of Gene Mapping Membranes, K. Rudd for communication of unpublished data, and W. Dowhan and M. Winkler for clones used as probes. We especially acknowledge P. Berget, R. Maurer, and the students of the Advanced Bacterial Genetics course taught at Cold Spring Harbor Laboratory for their help in developing and applying many of the transposition and mapping techniques used to isolate and analyze auxotrophs.

This work was supported in part by PHS grant GM35247 to G.M.W.

#### REFERENCES

1. Bachmann, B. Personal communication.
2. Bachmann, B. J. 1987. Derivations and genotypes of some mutant derivatives of *Escherichia coli* K-12, p. 1190-1219. In F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
3. Berget, P. Personal communication.
4. Bio-Rad Laboratories. CHEF-DR11 instruction manual, appendix. Bio-Rad Laboratories, Richmond, Calif.
5. Bouché, J., S. Béjar, and K. Cam. 1990. Cooption of prophage genes: new data on the Kim prophage region of the *Escherichia coli* chromosome, p. 373-378. In K. Drlica and M. Riley (ed.), *The bacterial chromosome*. American Society for Microbiology, Washington, D.C.
6. Casadaban, M. J. 1976. Transposition and fusion of the *lac* genes to selected promoters in *Escherichia coli* using bacteriophage lambda and Mu. *J. Mol. Biol.* 104:541-555.
7. Chu, G., D. Vollrath, and R. W. Davis. 1986. Separation of large DNA molecules by contour-clamped homogeneous electric fields. *Science* 234:1582-1585.
8. Clarke, L. and J. Carbon. 1979. Selection of specific clones from colony banks by suppression or complementation tests. *Methods Enzymol.* 68:398-408.
9. Clowes, R. C., and W. Hayes. 1968. Experiments in microbial genetics. Blackwell Scientific Publications, Oxford.
10. Condemine, G., and C. L. Smith. 1990. Genetic mapping using large-DNA technology: alignment of *SfiI* and *AvrII* sites with the *NotI* genomic restriction map of *Escherichia coli* K-12, p. 53-60. In K. Drlica and M. Riley (ed.), *The bacterial chromosome*. American Society for Microbiology, Washington, D.C.
11. Davis, R. W., D. Botstein, and J. R. Roth. 1980. Advanced bacterial genetics: a manual for genetic engineering. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
12. Devereux, J., P. Haeblerli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* 12:387-395.
13. Elliott, T., and J. R. Roth. 1988. Characterization of Tn10d-Cam: a transposition-defective Tn10 specifying chloramphenicol resistance. *Mol. Gen. Genet.* 213:332-338.
14. Friedman, A. M., S. R. Long, S. E. Brown, W. J. Buikema, and F. M. Ausubel. 1982. Construction of a broad host range cosmid

- cloning vector and its use in the genetic analysis of *Rhizobium* mutants. *Gene* **18**:289–296.
15. Graña, D., T. Gardella, and M. M. Susskind. 1988. The effects of mutations in the *ant* promoter of phage P22 depend on context. *Genetics* **120**:319–327.
  16. Guyer, M. S., R. R. Reed, J. A. Steitz, and K. B. Low. 1981. Identification of a sex-factor-affinity site in *E. coli* as  $\gamma\delta$ . Cold Spring Harbor Symp. Quant. Biol. **45**:135–140.
  17. Hill, C. W., and B. W. Harnish. 1981. Inversions between ribosomal RNA genes of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **78**:7069–7072.
  18. Icho, T. 1988. Membrane-bound phosphatases in *Escherichia coli*: sequence of the *pgpB* gene and dual subcellular localization of the *pgpB* product. *J. Bacteriol.* **170**:5117–5124.
  19. Kohara, Y. 1990. Correlation between the physical and genetic maps of the *Escherichia coli* K-12 chromosome, p. 29–42. In K. Drlica and M. Riley (ed.), *The bacterial chromosome*. American Society for Microbiology, Washington, D.C.
  20. Kohara, Y., K. Akiyama, and K. Isono. 1987. The physical map of the whole *Escherichia coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. *Cell* **50**:495–508.
  21. Noda, A., J. B. Courtwright, P. F. Denor, G. Webb, Y. Kohara, and A. Ishihama. 1991. Rapid identification of specific genes in *E. coli* by hybridization to membranes containing the ordered set of phage clones. *BioTechniques* **10**:474–477.
  22. Rudd, K. E., W. Miller, J. Ostell, and D. A. Benson. 1990. Alignment of *Escherichia coli* K12 DNA sequences to a genomic restriction map. *Nucleic Acids Res.* **18**:313–321.
  23. Rudd, K. E., W. Miller, C. Werner, J. Ostell, C. Tolstoshev, and S. G. Satterfield. 1991. Mapping sequenced *E. coli* genes by computer: software, strategies, and examples. *Nucleic Acids Res.* **19**:637–647.
  24. Schweizer, H. P., and P. Datta. 1990. Physical map location of the *tdc* operon of *Escherichia coli*. *J. Bacteriol.* **172**:2825.
  25. Silhavy, T. J., M. L. Berman and L. W. Enquist. 1984. Experiments with gene fusions. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
  26. Singer, M., T. A. Baker, G. Schnitzler, S. M. Deischel, M. Goel, W. Dove, K. J. Jaacks, A. D. Grossman, J. W. Erickson, and C. A. Gross. 1989. A collection of strains containing genetically linked alternating antibiotic resistance elements for genetic mapping of *Escherichia coli*. *Microbiol. Rev.* **53**:1–24.
  27. Smith, C. L., J. G. Econome, A. Schutt, S. Klco, and C. R. Cantor. 1987. A physical map of the *Escherichia coli* K12 genome. *Science* **236**:1448–1453.
  28. Tabata, S., A. Higashitani, M. Takanami, K. Akiyama, Y. Kohara, Y. Nishimura, A. Nishimura, S. Yasuda, and Y. Hirota. 1989. Construction of an ordered cosmid collection of the *Escherichia coli* K-12 W3110 chromosome. *J. Bacteriol.* **171**:1214–1218.
  29. Winkler, M. Personal communication.